Assessing the effects of a sequestered germline on interdomain lateral gene transfer in Metazoa

Lindy Jensen[1§], Jessica R. Grant[1], H. Dail Laughinghouse IV[1], and Laura A. Katz[1,2]*

[1] Department of Biological Sciences, Smith College, Northampton, Massachusetts 01063, USA

[2] Program in Organismic and Evolutionary Biology, University of Massachusetts, Amherst, Massachusetts 01003, USA

[§]Present address: Department of Molecular & Integrative Physiology, University of Michigan, Ann Arbor, Michigan 48109, USA

*Author for correspondence: lkatz@smith.edu

Running title: Patterns of interdomain lateral gene transfer from bacteria and archaea to Metazoa and the sequestered germline

Keywords: Lateral Gene Transfer, Horizontal Gene Transfer, weak link model, gene loss, Metazoa, Animals

Word Count: 5510

18 Table Count (Supporting Information): 2 (2)

19 Figure Count (Supporting Information): 3 (1)

20 Data archival location: Dryad

ABSTRACT

A sequestered germline in Metazoa has been argued to be an obstacle to lateral gene transfer (LGT), though few studies have specifically assessed this claim. Here we test the hypothesis that the origin of a sequestered germline reduced LGT events in Bilateria (i.e. triploblast lineages) as compared to early-diverging Metazoa (i.e. Ctenophora, Cnidaria, Porifera, and Placozoa). We analyze single-gene phylogenies generated with over 900 species, sampled from among Bacteria, Archaea and Eukaryota to identify well-supported interdomain LGTs. We focus on interdomain LGT (i.e. those between prokaryotes and Metazoa) as systematic errors in single gene tree reconstruction create uncertainties for interpreting eukaryote-to-eukaryotes transfers. The breadth of the sampled Metazoa enables us to estimate the timing of LGTs, and to examine the pattern before *vs.* after the evolution of a sequestered germline. We identified 58 LGTs found only in Metazoa and prokaryotes (i.e. bacteria and/or archaea), and 7 genes transferred from prokaryotes into Metazoa plus one other eukaryotic clade. Our analyses indicate that more interdomain transfers occurred before the development of a sequestered germline, consistent with the hypothesis that this feature is an obstacle to LGT.

INTRODUCTION

Lateral gene transfer (LGT), the exchange of genetic material from one lineage to another, has had a substantial impact on biological diversity on Earth (Dagan and Martin 2009, Davison 1999, Gogarten, Doolittle, and Lawrence 2002, Jain et al. 2003, Katz 2012, Soucy, Huang, and Gogarten 2015, Syvanen 2012). Though lateral gene transfers within and between bacteria and archaea are relatively common (Doolittle 1999, Grant and Katz 2014, Katz 2015, Nelson et al. 1999, Ochman, Lawrence, and Groisman 2000), the contributions of interdomain LGT from bacteria or archaea into eukaryotes is less well understood (Andersson 2005, Boto 2014, Keeling and Palmer 2008, Syvanen 2012, Wijayawardena, Minchella, and DeWoody 2013). In particular, the contribution of interdomain LGT across the metazoan tree of life is unclear due to a dearth of taxon-rich analyses that include both metazoan and bacterial/archaeal lineages.

There are numerous approaches for distinguishing laterally transferred genes from vertically inherited genes (i.e. genes passed down within a lineage). One approach is to analyze phylogenetic trees generated from homologs sampled from diverse taxa, and then examine the results for incongruence (Keeling and Palmer 2008). A gene tree is identified as incongruent when the evolutionary relationships between the species are discordant with established phylogenetic relationships and LGT provides a parsimonious explanation for the incongruence (Dagan 2011, Keeling and Palmer 2008). While this methodology is appropriate for identifying ancient LGT events, limited or biased taxon sampling and poor data quality can be problematic in interpreting patterns in single gene trees (Ragan, Harlow, and Beiko 2006). Fortunately, the growing availability of whole genomes is alleviating these shortcomings.

An alternative to gene tree analyses is to use methods that assess patterns in sequence

composition, though these methods work best for recent events. For example, differences in the composition of one gene compared to another or codon frequencies can be compared (Lawrence and Hendrickson 2003, Ragan 2001). While these methods often yield larger numbers of anomalies and don't require as broad a taxon sampling, these parametric methods have their own set of disadvantages as distinguishing causes for changes in compositional bias can be challenging.. Most importantly for the current study, foreign genes begin to resemble the genome around them over time and gene loss/duplication can confound results (Lawrence and Ochman 2002). As a result, studies examining the same transcriptomes but using different parameters to generate different sets of putative lateral gene transfers can yield conflicting results (Ragan, Harlow, and Beiko 2006).

After finding laterally transferred genes, determining the function of these genes can reveal the role of LGT within lineages. Laterally transferred genes arguably confer adaptive functions and novel traits (Schönknecht et al. 2013), the most notable example being the massive endosymbiotic gene transfers (EGT) in eukaryotes from mitochondria and plastids (Keeling and Palmer 2008, Andersson 2005, Katz 2002). However, outside of EGT, laterally transferred genes found in eukaryotes tend to be metabolic genes (Jain, Rivera, and Lake 1999, Jain et al. 2003, Keeling and Palmer 2008, Koning et al. 2000). Similarly, in bacteria many LGTs are involved in metabolic functions (Pál, Papp, and Lercher 2005, Lehmann and Richardson 2010, Whitaker, McConkey, and Westhead 2009). Dubbed "The Complexity Hypothesis", Jain et al. (1999) proposed that the 'operational' genes often involved in cellular metabolism are much more likely to be successfully transferred than 'informational' genes that are part of complex networks. In other words, metabolic and house-keeping processes that only involve two or three protein interactions are more likely to be successfully exchanged among lineages (Whitaker, McConkey, and Westhead 2009).

The impact of LGT in Metazoa has been assumed to be low based on the hypothesis

that the sequestered metazoan germline is an obstacle to LGT (Andersson 2005, Andersson, Doolittle, and Nesbø 2001, Boto 2010, Doolittle 1999, Huang 2013, Katz 2002, Keeling and Palmer 2008, Schönknecht, Weber, and Lercher 2014). However, not all Metazoa have a sequestered germline, as germline determination is more fluid among early-diverging lineages (i.e. Ctenophora, Cnidaria, Porifera, and Placozoa). The absence of a sequestered germline in early-diverging Metazoa might lead to an increase in LGT rates because: 1) germ cells are more accessible to foreign DNA throughout their lifespan; and 2) the presence of totipotent cell lines that can take on a variety of fates during the life cycle (including gametes) provides more opportunities for transfer (Alie et al. 2011, David 2012, Funayama 2010, Juliano, Swartz, and Wessel 2010). Germ cells of some early-diverging metazoan lineages often move throughout the body or line the inside of the gut (Brusca and Brusca 2003, Gaino, Burlando, and Buffa 1986, Gaino et al. 1984, Nishimiya-Fujisawa and Kobayashi 2012, Simpson 1980, 1984). In contrast, germ cells in Bilateria (i.e. triploblasts) generally migrate during development to a specific location away from the gut where they are sheltered from foreign DNA by layers of tissue (Ewen-Campen, Schwager, and Extavour 2010, Extavour 2007, Lehmann and Richardson 2010, Starz-Gaiano and Lehmann 2001). Among Bilateria, it is generally only the primordial germ cells (PGCs) that are capable of differentiating into gametes, and somatic cells have no gametogenic potential (Extavour and Akam 2003).

Numerous studies have explored laterally transferred genes in metazoan lineages demonstrating that LGT is at least possible in lineages with and without sequestered germline (Boschetti et al. 2012, Cobbs et al. 2013, Crisp et al. 2015, Danchin et al. 2010, Gladyshev, Meselson, and Arkhipova 2008, Klasson et al. 2009, Kondrashov et al. 2006, Nikoh and Nakabachi 2009, Paganini et al. 2012, Sloan and Moran 2012, Starcevic et al. 2008). Perhaps

the most well-known example stems from the symbiotic relationship involving arthropods and nematodes and the bacterial genus *Wolbachia*, which has been shown to be a donor of laterally transferred genes in diverse insect lineages (Dunning Hotopp 2011). In nematodes, lateral gene transfer may have played a key role in the development of plant parasitism (Haegeman, Jones, and Danchin 2011). Laterally transferred genes have also been found in chordates as the potential source for a cellulose synthase gene in *Ciona intestinalis* (Nakashima et al. 2004). Among early-diverging Metazoa, the biomineralization gene in sponges may be the product of LGT (Jackson et al. 2011). Lastly, 71 genes found in the detailed examination of the *Hydra* genome had strongly supported LGT origins (Chapman et al. 2010). These studies primarily focus on detecting very recent LGT in a small number of metazoan species.

Here we assess a set of single gene trees generated by a taxon-rich phylogenomic pipeline (Grant and Katz 2014) to test the hypothesis that the evolution of a sequestered germline reduced the number of lateral gene transfers into Bilateria. In order to diminish the chances of misinterpreting the systematic errors that confound the topologies of individual gene trees at the time scales relevant for this study (i.e. 100s of millions of years), we chose to focus on interdomain 'presence/absence' genes. These genes are defined based on their presence only in a monophyletic clade of Metazoa plus prokaryotes (bacteria and/or archaea), and their absence in all other eukaryotes (Figure 1). The resulting 'presence/absence' genes were either a) present in the last common ancestor of Metazoa and bacteria or archaea, then lost in every other lineage since, or b) the product of a lateral gene transfer event. Placing the resulting LGTs onto a metazoan phylogeny allows us to determine the approximate time of lateral transfer for each individual gene. As the evolution of a sequestered germline happened near the time protostomes and deuterostomes diverged (i.e. just at the base of Bilateria), we

compared the number of laterally transferred genes in Bilateria to the number of genes laterally transferred in early-diverging lineages that lack a sequestered germline.

METHODS

Phylogenomic Pipeline

Single gene alignments and trees were built using a phylogenetic pipeline (Grant and Katz 2014) with a modified taxonomic dataset that aimed to capture a greater number of animal lineages. In brief, this pipeline starts with set of protein coding genes from taxa of interest, and then uses BLAST (Altschul et al. 1990) to compile them into homologous clusters as determined by OrthoMCL DB (Chen et al. 2006). Then, custom python scripts refined data from each taxon by removing nearly identical sequences (i.e. alleles) and highly divergent sequences (i.e. those sharing only motifs) as described in Grant and Katz (2014). Multisequence alignments are generated using Guidance (Penn et al. 2010), which also allows for the automated removal of poorly-fitting taxa and columns.

The pipeline includes data from 910 taxa chosen to broadly represent the three domains of life, including 489 eukaryotes, 303 bacteria and 118 archaea. The metazoan taxa were chosen to capture a broad diversity of lineages in order to accurately place ancient transfer events in metazoan evolution. At the same time, we sought some level of evenness between the early diverging lineages and the more thoroughly sampled bilaterian lineages to prevent bias from oversampling recent events in one portion of the tree compared to the other. In total, 61 Metazoa were included in this study, of which 25 were early-diverging and 36 were Bilaterian (19 protostomes and 17 deuterostomes; Table S1). The level of available data varied among lineages with some represented by whole genomes and others by ESTs or

RNAseq data (Table S1).

Finding presence/absence genes in only Metazoa and bacteria/archaea

We identified a total of 58 genes that are found only in Metazoa and bacteria and/or archaea (i.e. presence/absence data; Figure 1) from the 1575 alignments that included >2 Metazoa taxa and >2 bacterial and/or archaeal lineages (please see Supplementary Figure S1 for a schematic of the process). With the original 1575 trees as a starting point, we used custom Python scripts based on the tree walking methods in p4 (Price, Dehal, and Arkin 2009, 2010) to assess the relationships in the tree topologies and identify 310 candidate laterally transferred genes (i.e. genes with Metazoa sister to bacteria and/or archaea). To assess presence/absence data further, we searched that pool of 310 candidates for genes found only in Metazoa and bacteria and/or archaea using additional custom Python scripts (Figure 1). To control for contamination and recent LGT, each gene tree could contain up to two non-sister (e.g. not next to one another) single eukaryotes, resulting in 193 potential presence/absence genes. The pipeline, which was developed to assess deep phylogenetic relationships, uses a relaxed criterion for sequence and taxon scoring in Guidance (Penn et al. 2010). To increase our confidence in the homology of the presence/absence genes, we realigned the 193 gene sequences with the more stringent default settings for removing non-orthologous taxa. This generated a set of 72 putative presence/absence genes. We then eliminated any alignments with fewer than 10 total sequences or only 1 metazoan species (regardless of the number of paralogs). The remaining 65 alignments were then inspected by eye and an additional six alignments were removed because the alignment quality did not meet our strict standards, for example by having non-monophyletic Metazoa. This resulted in a final count of 58 presence/absence genes.

Topologies with two LGT Events

In addition to the presence/absence genes, we also found seven genes with topologies indicative of two interdomain LGT events: one into Metazoa and another into a different eukaryotic clade. Among the remaining 1517 single-gene trees after finding presence/absence genes, 15 gene trees showed topologies consistent with two interdomain LGT events involving Metazoa. We further used RAxML, (PROTGAMMALG model with default parameters, 100 bootstrap replicates) to bootstrap these trees and assess the support for the topological relationships. We found two topological trends: a metazoan clade and one other eukaryotic clade each nested within bacteria (Figure 2a), and a metazoan clade and another clade of two photosynthetic eukaryotic groups, one nested within the other (Figure 2b). We required the Metazoa and other eukaryotic clade to be separated by a bootstrap value of 70 or greater to increase confidence in resulting data. These criteria resulted in detection of seven genes, each with two topologically well-supported lateral gene transfer events.

Determining the Timing of LGT Events

To determine the approximate time of transfer and placement on our phylogeny, we assumed that each gene was transferred into Metazoa once and into the last common ancestors of the taxa containing the gene. Any nested 'missing' taxa were assumed to have lost the gene, which is consistent with evidence for rampant gene loss among eukaryotes (Katz 2015, Wolf and Koonin 2013). For example, if the gene was found in cnidarian and bilaterian taxa, we inferred the LGT event occurred before the divergence of cnidarians and bilaterians (Figure 3; event D). An alternative here would be to infer one interdomain LGT followed by LGT within Metazoa, which is possible but less likely. If a gene was only found

204 in cnidarians, the transfer was assumed to have occurred after the divergence of cnidarians
205 (Figure 3; event E). We recognize that our approach will misrepresent the timing of LGT
206 events that occurred early and were then lost in non-triploblastic lineages.
207 For the purpose of visualizing the approximate transfer times for the genes we found,
208 we created the synthetic tree shown in Figure 3. We generated this phylogeny of the
209 metazoan taxa included in our pipeline by concatenating vertically transferred genes and
210 building a phylogeny in RAxML. Then, the tree was adjusted to reflect the arrangement of
211 major metazoan phyla found in (Ryan et al. 2013), creating a synthetic tree of relationships
212 among taxa in our pipeline. To illustrate the approximate timing of events, branches with an
213 LGT event was assigned a letter (Figure 3) corresponding to Table 1 and Table S2.
214
215 *Assessing Gene Function*
216 A representative sequence from each of the 58 presence/absence genes was passed
217 through the KEGG BLAST tool (Kanehisa and Goto 2000) to determine gene names and
218 functions based upon the KEGG BRITE Hierarchy Ontology. We used a human or other
219 'model' organism as the representative sequence for a gene when available. If multiple
220 functional categories were available, we choose the function that appeared most applicable to
221 both bacteria/archaea and metazoan, though certainly functions have changed over time.
222
223
224 **RESULTS**
225 Consistent with the hypothesis that a sequestered germline creates an obstacle to LGT
226 in Metazoa (e.g. Andersson 2005, Keeling and Palmer 2008), our analyses reveal a greater

number of gene transfers before the development of a sequestered germline in bilaterian animals. A total of 58 genes satisfied our 'presence/absence' criteria for LGT as they are only found in bacteria/archaea and Metazoa (Table 2, Figure 1). Of those 58 interdomain LGTs, we found 34 genes transferred prior to the evolution of the sequestered germline (Table 2, Figure 3; events A-F) as these genes were present in at least one of the 25 early-diverging lineages included in our analyses (Table S1). In contrast, we only detect 24 LGT events restricted to Bilateria (Table 2, Figure 3; events G-P), where we had 36 species sampled (Table S1). If the detection of LGT events is dependent only on the number of species sampled, then the distributions – 34 LGTs into 25 species and 24 LGTs into 36 species – are significantly different based on goodness-of-fit G-test (p-value = 0.0069), with an over-abundance of events into early-diverging lineages.

Only one gene (OG5_139285, *RENBP*) appears to have been transferred from prokaryotes into the last common ancestor of Metazoa (Table 1, Table S2, Figure 3; event A). Of the remaining 33 genes, our analyses reveal that seven genes were transferred before the divergence of sponges (Table 2, Figure 3; event B), four before the divergence of Placozoa (Table 2, Figure 3; event C), 13 before the divergence of cnidarians (Table 2, Figure 3; event D), and seven before the divergence of protostomes and deuterostomes (Table 2, Figure 3; event F); though interpretation must be made with caution given debates on relationships among these early lineages (Shu et al. 2014, Just, Kristensen, and Olesen 2014, Ryan et al. 2013, Edgecombe et al. 2011). Finally, two laterally transferred genes were found in cnidarians alone (Figure 3; event E).

Within Bilateria, we found a total of 24 genes laterally transferred into either protostomes or deuterostomes and thus after acquisition of a sequestered germline at the base of Bilateria. Ten of these genes were found in only deuterostomes, while the other 14 were

found only in protostomes (Table 2). Further, seven of the ten were present only in vertebrates (fishes, mammals, and birds; Figure 3; event H) and two genes were present exclusively in mammals (Figure 3; event I). Of the 14 genes laterally transferred into protostomes, only one gene was transferred before the split of Lophotrocozoa and Ecdysozoa (Figure 3; event J); the remaining 13 were present exclusively in Ecdysozoa. Nine of those 13 were transferred into arthropods (Figure 3; events M-P). Two genes were present in just nematodes (Figure 3; event L) and two genes appear to have been transferred before the split of nematodes and arthropods (Figure 3; event K).

Out of the seven gene topologies supporting two LGT events (i.e. genes present in bacteria/archaea, Metazoa and one other eukaryotic group), all seven were transferred before the development of a sequestered germline. Two trends within the topologies emerged upon examination: the non-metazoan monophyletic branch contained eukaryotes of a single major clade (often parasitic, e.g. *Entamoeba*, and *Trichomonas vaginalis*) (Figure 2a) or consisted of two plastid-containing eukaryotes (one nested within the other, such as haptophytes and photosynthetic dinoflagellates) (Figure 2b). The trees with transfers into plastid-containing eukaryotes suggest that these genes may be the result of endosymbiotic gene transfer into the photosynthetic lineages.

Laterally Transferred Genes Serve Primarily Metabolic Functions

We used KEGG classification to assign putative gene functions for 34 of the 58 genes, with the remaining gene functions unknown/hypothetical. Of the 34 genes with known functions, the majority (16) are metabolic (Table 1, Table S2). Transporters made up the next most numerous category with 13 genes, eight of which belonged to the solute carrier family (SLC; Table 1). The remaining genes represent diverse functions: DNA repair (2), genetic

information processing (1), spliceosome-associated (1), and signal transduction (1) (Table 1). The large number of unknown/hypothetic genes may represent genes with lineage-specific functions.

DISCUSSION

Estimated rates of LGT are affected by germline sequestration, gene loss, genome size, and data quality

Our data reveal that the evolution of a sequestered germline is associated with a decrease in the rate of interdomain lateral gene transfer (Figure 3, Table 2). We find a greater proportion of interdomain LGTs (34 out of 58) into lineages that predate the evolution of a sequestered germline even though we sampled fewer of these lineages than Bilateria (25 vs 36 species, respectively; Fig 1). Additional biases in our data include a greater number and higher quality data (i.e. more complete genomes) from Bilateria, and both a longer timespan for gene loss and tendency for smaller genomes in early-diverging lineages. Despite these, we found that the number of genes transferred into Metazoa before the evolution of a sequestered germline is significantly different (p-value <0.01) than the number of genes transferred into metazoans with a sequestered germline.

Previous claims that LGT is rare in Metazoa have generally failed to consider the fact that not all metazoans have a sequestered germline. Early-diverging metazoan lineages generally lack a sequestered germline and there is a greater chance for their gamete-producing cells to come in contact with foreign DNA. For example, cnidarian gonads are most often located in the gastrodermis, the interior lining of the cnidarian ‘gut’ or other structures close to the gut (Brusca and Brusca 2003, Eckelbarger, Tyler, and Langton 1998, Galliot et al. 2006, Nishimiya-Fujisawa and Kobayashi 2012,

298 Wedi and Dunn 1983). In ctenophores, the gonads form along different ‘canals’ depending upon the
299 species, all of which connect to the ‘gut’ at some point (Brusca and Brusca 2003, Harbison and Miller
300 1986, Komai 1922). Although poriferans have no designated gonads, two cell types are typically
301 responsible for transforming into gametes: 1) the choanocytes that line the internal water-filtering
302 systems and are responsible for catching and ingesting nutrients and 2) archeocytes, a totipotent
303 amoeboid cell that moves throughout the body of the sponge (Brusca and Brusca 2003, Funayama
304 2010, Gaino, Burlando, and Buffa 1986, Gaino et al. 1984, Simpson 1980, 1984, Tsurumi and
305 Reiswig 1997). As such, the gametogenic cells in these early-diverging lineages either exist in close
306 quarters to digestion, an area likely to come in contact with foreign DNA, or are responsible for
307 ingestion themselves, all of which could increase their rate of LGT in comparison to Bilateria
308 lineages.

309 An alternative explanation for the larger proportion of genes transferred into early-diverging
310 lineages is that early-diverging Metazoa have been around longer and thus have had more time for
311 LGT to occur. Although the exact dating and divergence of Metazoa is debated (Shu et al. 2014,
312 Edgecombe et al. 2011), Metazoa likely first emerged around 800 million years ago (mya) while the
313 divergence of Bilateria occurred approximately 550 mya (Erwin et al. 2011), leaving ~250 mya for
314 early-diverging metazoans to acquire more genes from bacteria (as well as an extra ~250 million years
315 to lose genes; see next paragraph). However, we sampled more Bilaterian than early-diverging
316 lineages and the species richness of Bilaterian lineages (particularly Arthropoda) dwarf that of early-
317 diverging metazoans (Zhang 2013). Both these factors should bias towards more detection of LGT
318 among Bilateria, even given our requirement that an LGT event be found in at least two descendant
319 lineages. Instead, our analyses show that a larger number of genes were transferred into Metazoa
320 during the smaller timespan before the evolution of a sequestered germline, whereas we found fewer
321 genes transferred into Bilateria since the development of a germline. Our data reveal that at least 32
322 genes were transferred from prokaryotes in the relatively short ~250 million years between the

divergence of Metazoa and the development of a sequestered germline (A + B + C + D + F, all transfers that were present in Early Diverging and Bilaterian taxa), while only 24 genes were transferred into Bilateria in the ~550 million years since.

An additional caveat that may lead to an underestimate of early LGT events is the relatively small genomes in the sampled early-diverging metazoan lineages. For example, eleven of the deuterostome genomes included in this study have one billion or more base pairs compared to just two of the protostomes and two of the early-diverging lineages (Gregory 2016). Given that gene loss is so common throughout evolution (Keeling and Palmer 2008, Katz 2015, Wolf and Koonin 2013), it's possible that many LGTs are lost during genomic streamlining in metazoans with very small genomes, like those in early-diverging species. Conversely, the larger genomes found throughout Bilateria might create the opposite problem, particularly in the deuterostomes: larger genomes may undergo less genomic streamlining (Gregory 2001, Lynch and Conery 2000, Wolf and Koonin 2013), resulting in more LGT-tolerant genomes (Table S1) and possibly an overestimate of LGTs. For example, the apparent pulse of transfer events in vertebrates (Table 2, Figure 3; event H) could be explained as either independent LGTs into deuterostomes, or the retention of genes transferred sometime before the divergence of deuterostomes and protostomes but then lost in lineages with smaller genomes.

Finally, the quality of the data also affects the patterns found in this study even though we used strict criteria for interpreting interdomain LGTs. Many of the early-diverging taxa do not yet have whole genome sequences available and, as a result, we used ESTs or RNAseq data. As such, genes transferred before the development of a sequestered germline may not have been present in these incomplete datasets. For example, the 13 genes in event D (Figure 3) may have been transferred sometime before the divergence of cnidarians, but may not have been detected in our analyses due to poor sequencing data from early-diverging lineages. Overall, the quality and quantity of Bilaterian

347 genomic data are better than for early diverging lineages, which makes finding laterally transferred
348 genes in early diverging lineages more challenging.

349 When combined with our strict criteria (i.e. 'presence/absence' genes found only in
350 prokaryotes and at least two metazoan), the differences in the quality of the underlying data likely
351 lowered the number of inferred LGTs found among individual early-diverging phyla (e.g. Figure 3;
352 event E). For example, our criteria require that the gene be present in at least two Metazoa, but since
353 sequences are only available for one species (*Trichoplax adhaerens*) in phylum Placozoa, we cannot
354 detect any ancient LGT that occurred after their divergence. In sum, the biased distribution of non-
355 whole genome data among early-diverging taxa likely leads to an underestimate of the role of a
356 sequestered germline in blocking LGTs as early LGT events were more likely to go undetected.

357

358 *Life history could result in "weak-links" and higher chances of LGT among different lineages*

359 Lateral gene transfers might also be enhanced in lineages that have life history stages that
360 increase exposure to potential donor lineages (i.e. the "weak-link model" (Huang 2013)). Weak-links
361 associated with specific lifestyles could include harboring endosymbionts or parasites (e.g. *Wolbachia*
362 in arthropods and nematodes), external fertilization, and the retention of totipotent cells throughout
363 adulthood (often associated with regeneration and asexual reproduction) (Huang 2013). Many early-
364 diverging Metazoa exhibit these traits (David 2012, Extavour et al. 2005, Funayama 2010, Galliot et
365 al. 2006, Juliano, Swartz, and Wessel 2010, Nishimiya-Fujisawa and Kobayashi 2012, Tsurumi and
366 Reiswig 1997), but some Bilateria do as well including Planaria (Agata and Inoue 2012, Newmark,
367 Wang, and Chong 2008), some species of newt (Okamoto et al. 2007), echinoderms (Shibata et al.
368 2010), and others (Collins et al. 2013, Giani et al. 2011, Newmark, Wang, and Chong 2008). In
369 particular, bdelloid rotifers serve as an exceptional example of a 'weak-link' lifestyle leading to
370 rampant LGT in a protostome; bdelloid rotifers survive desiccation and genome decay periodically,

371 which likely makes incorporating laterally transferred genes into the genome much easier (Flot et al.
372 2013). Indeed, recent studies examining several species of bdelloid rotifers have found that large
373 portions of their genomes are from foreign sources (Boschetti et al. 2012, Flot et al. 2013, Eyres et al.
374 2015). Although we did not include any bdelloid rotifers in our study, we do agree that bdelloid
375 rotifers serve as a clear example of 'weak-link' taxa with increased potential for LGT. We did sample
376 two bilaterian species (*Schistosoma mansoni* and *Capitella teleta*) exhibiting weak-link traits, and we
377 find no evidence of increased LGTs into these lineages (Figure 3), possibly because we did not
378 include LGTs into a single species. However, we predict that additional examples will be found with
379 the emergence of data from more diverse lineages.

380

381 *Comparisons of our insights to other studies*

382 We also used our pipeline to examine laterally transferred genes reported in other studies
383 (Boschetti et al. 2012, Crisp et al. 2015). Upon examining 394 genes individually reported to be found
384 in *H. sapiens*, *D. melanogaster*, and *C. elegans*, we found that nearly half (177) of these genes
385 mapped to the same Ortholog Group (i.e. they represent paralogs) number, leaving 217 putative
386 laterally transferred genes. Of these, 97 gene trees were found in our pipeline and, using custom
387 Python scripts, we asked to what degree the remaining 97 gene trees satisfied our criteria for LGT
388 between bacteria/archaea and metazoan . Only three genes passed our criteria for interdomain lateral
389 gene transfer and are included and reported in this study. Two are part of the 58 presence/absence
390 genes (Figure 3 and Table S2; event L), while one is in the reported topologies with more than one
391 LGT (a gene named NQO1, transferred before deuterostomes and protostomes diverged,
392 OG5_138113). The remaining genes did not match our more conservative criteria as they are either
393 not found in more than one metazoan 'recipient' lineage OR are found in gene trees with other
394 eukaryotes; we choose to focus only on interdomain LGTs given the systematic errors that can lead to
395 incorrect topologies in single gene trees capturing hundreds of millions of years of evolution.

396 The small overlap between our final dataset and the datasets of Crisp et al. (2015) and Boschetti
397 et al. (2012) is expected. It's common for studies that use different methodologies and criteria to come
398 up with different final datasets (Lawrence and Hendrickson 2003, Ragan 2001, Dagan 2011,
399 Lawrence and Ochman 2002). Also, since these studies considered intradomain LGT, many of the
400 genes found in Crisp et al. (2015) and Boschetti et al. (2012) are also present in non-metazoan
401 eukaryotes and would have been removed under our criteria for interdomain LGTs. Furthermore, both
402 Boschetti et al. (2012) and Crisp et al. (2015) used surrogate methods with unique parameters to
403 search for LGTs in *C. elegans* and reported different numbers of found genes between themselves.
404 Ultimately, all three studies (Boschetti et al. (2012), Crisp et al. (2015), and our own) used different
405 approaches and parameters to address distinct questions, making it difficult to accurately compare
406 across the three.

407

408 *The majority of proteins involved in LGT events have metabolic or transport roles*

409 Although only 34 of the 58 presence/absence genes had known functions, the majority of these
410 (16) are involved in metabolism, with the remaining 18 genes scattered into a few other functional
411 categories (Table 1). This is consistent with the types of genes transferred among bacteria (Pál, Papp,
412 and Lercher 2005). While LGT has been considered a source of adaptive traits (Schönknecht, Weber,
413 and Lercher 2014), the majority of LGTs tend to serve metabolic functions, perhaps because these
414 genes are able to integrate into comparatively simple metabolic gene networks as opposed to more
415 complicated ones (Jain, Rivera, and Lake 1999). Transporters were the next most numerous (13
416 genes) functional category found, likely because membrane transport genes are quite common.
417 Membrane proteins themselves make up 27% of all human protein-coding genes, of which the solute
418 carrier family is the second largest family with at least 50 defined families (Fredriksson et al. 2008,
419 Almen et al. 2009). The eight SLC genes found here (Table 1, Table S2) were from a variety of SLC
420 families.

Limitations and future work

The methodology we used in this study enabled us to find instances of interdomain LGT in lineages across the metazoan tree of life. However, we acknowledge that there are numerous limitations to our analyses. For example, some taxa have only transcriptome data or limited amounts of whole genome data available, and these biases are likely greater among early-diverging lineages. Moreover, our inferences are dependent on our taxon sampling. We carefully assembled a sampling of Metazoa that balanced phylogenetic breadth and evenness; there are many more triploblast lineages that we could have sampled and inclusion of these would surely increase the number of recent LGT events inferred. However, early-diverging lineages lack of similar levels of sampling and so creating additional unevenness would introduce further bias. Future studies with additional taxa will surely change the details of inferences on LGTs as increasing taxon sampling from lineages across the metazoan tree of life will increase the power of the analyses, allowing future studies to more easily assess the pattern of interdomain LGTs over time.

REFERENCES

Agata, K., and T. Inoue. 2012. "Survey of the differences between regenerative and non-regenerative animals." *Development Growth & Differentiation* no. 54 (2):143-152. doi: 10.1111/j.1440-169X.2011.01323.x.

Alie, A., L. Leclere, M. Jager, C. Dayraud, P. R. Chang, H. Le Guyader, E. Queinnec, and M. Manuel. 2011. "Somatic stem cells express Piwi and Vasa genes in an adult ctenophore: Ancient association of "germline genes" with stemness." *Developmental Biology* no. 350 (1):183-197. doi: 10.1016/j.ydbio.2010.10.019.

Almen, M. S., K. J. V. Nordstrom, R. Fredriksson, and H. B. Schioth. 2009. "Mapping the human membrane proteome: a majority of the human membrane proteins can be classified according to function and evolutionary origin." *Bmc Biology* no. 7:14. doi: 10.1186/1741-7007-7-50.

Altschul, Stephen F., Warren Gish, Webb Miller, Eugene W. Myers, and David J. Lipman. 1990. "Basic local alignment search tool." *Journal of molecular biology* no. 215 (3):403-410.

Andersson, J. O. 2005. "Lateral gene transfer in eukaryotes." *Cellular and Molecular Life Sciences CMLS* no. 62 (11):1182-1197. doi: 10.1007/s00018-005-4539-z.

Andersson, Jan O., W. Ford Doolittle, and Camilla L. Nesbø. 2001. "Are there bugs in our genome?" *Science* no. 292 (5523):1848-1850.

Boschetti, Chiara, Adrian Carr, Alastair Crisp, Isobel Eyres, Yuan Wang-Koh, Esther Lubzens, Timothy G. Barraclough, Gos Micklem, and Alan Tunnacliffe. 2012. "Biochemical diversification through foreign gene expression in bdelloid rotifers." *PLoS Genet* no. 8 (11):e1003035. doi: 10.1371/journal.pgen.1003035.

457 Boto, Luis. 2010. "Horizontal gene transfer in evolution: facts and challenges." *Proceedings of the*
458 *Royal Society B: Biological Sciences* no. 277 (1683):819-827. doi: 10.1098/rspb.2009.1679.

459 Boto, Luis. 2014. Horizontal gene transfer in the acquisition of novel traits by metazoans. Paper read
460 at Proc. R. Soc. B.

461 Brusca, R. C., and G. J. Brusca. 2003. *Invertebrates*. Sunderland: Sinauer.

462 Chapman, Jarrod A., Ewen F. Kirkness, Oleg Simakov, Steven E. Hampson, Therese Mitros, Thomas
463 Weinmaier, Thomas Rattei, Prakash G. Balasubramanian, Jon Borman, Dana Busam, Kathryn
464 Disbennett, Cynthia Pfannkoch, Nadezhda Sumin, Granger G. Sutton, Lakshmi Devi
465 Viswanathan, Brian Walenz, David M. Goodstein, Uffe Hellsten, Takeshi Kawashima, Simon
466 E. Prochnik, Nicholas H. Putnam, Shengquiang Shu, Bruce Blumberg, Catherine E. Dana,
467 Lydia Gee, Dennis F. Kibler, Lee Law, Dirk Lindgens, Daniel E. Martinez, Jisong Peng,
468 Philip A. Wigge, Bianca Bertulat, Corina Guder, Yukio Nakamura, Suat Ozbek, Hiroshi
469 Watanabe, Konstantin Khalturin, Georg Hemmrich, André Franke, René Augustin, Sebastian
470 Fraune, Eisuke Hayakawa, Shiho Hayakawa, Mamiko Hirose, Jung Shan Hwang, Kazuho
471 Ikeo, Chiemi Nishimiya-Fujisawa, Atshushi Ogura, Toshio Takahashi, Patrick R. H.
472 Steinmetz, Xiaoming Zhang, Roland Aufschnaiter, Marie-Kristin Eder, Anne-Kathrin Gorny,
473 Willi Salvenmoser, Alysha M. Heimberg, Benjamin M. Wheeler, Kevin J. Peterson, Angelika
474 Böttger, Patrick Tischler, Alexander Wolf, Takashi Gojobori, Karin A. Remington, Robert L.
475 Strausberg, J. Craig Venter, Ulrich Technau, Bert Hobmayer, Thomas C. G. Bosch, Thomas
476 W. Holstein, Toshitaka Fujisawa, Hans R. Bode, Charles N. David, Daniel S. Rokhsar, and
477 Robert E. Steele. 2010. "The dynamic genome of *Hydra*." *Nature* no. 464 (7288):592-596.
478 doi: 10.1038/nature08830.

479 Chen, Feng, Aaron J. Mackey, Christian J. Stoeckert, and David S. Roos. 2006. "OrthoMCL-DB:
480 querying a comprehensive multi-species collection of ortholog groups." *Nucleic acids*
481 *research* no. 34 (suppl 1):D363-D368.

482 Cobbs, Cassidy, Jeremy Heath, John O. Stireman, and Patrick Abbot. 2013. "Carotenoids in
483 unexpected places: gall midges, lateral gene transfer, and carotenoid biosynthesis in animals."
484 *Molecular phylogenetics and evolution* no. 68 (2):221-228.

485 Collins, J. J., B. Wang, B. G. Lambrus, M. E. Tharp, H. Iyer, and P. A. Newmark. 2013. "Adult
486 somatic stem cells in the human parasite *Schistosoma mansoni*." *Nature* no. 494 (7438):476-
487 479. doi: 10.1038/nature11924.

488 Crisp, Alastair, Chiara Boschetti, Malcolm Perry, Alan Tunnacliffe, and Gos Micklem. 2015.
489 "Expression of multiple horizontally acquired genes is a hallmark of both vertebrate and
490 invertebrate genomes." *Genome biology* no. 16 (1):50. doi: 10.1186/s13059-015-0607-3.

491 Dagan, Tal. 2011. "Phylogenomic networks." *Trends in microbiology* no. 19 (10):483-491.

492 Dagan, Tal, and William Martin. 2009. "Getting a better picture of microbial evolution en route to a
493 network of genomes." *Philosophical Transaction: Biological Sciences* no. 364 (1527):21870-
494 2196.

495 Danchin, Etienne G. J., Marie-Noëlle Rosso, Paulo Vieira, Janice de Almeida-Engler, Pedro M.
496 Coutinho, Bernard Henrissat, and Pierre Abad. 2010. "Multiple lateral gene transfers and
497 duplications have promoted plant parasitism ability in nematodes." *Proceedings of the*
498 *National Academy of Sciences* no. 107 (41):17651-17656. doi: 10.1073/pnas.1008486107.

499 David, C. N. 2012. "Interstitial stem cells in *Hydra*: multipotency and decision-making." *International*
500 *Journal of Developmental Biology* no. 56 (6-8):489-497. doi: 10.1387/ijdb.113476cd.

501 Davison, John. 1999. "Genetic exchange between bacteria in the environment." *Plasmid* no. 42 (2):73-
502 91. doi: 10.1006/plas.1999.1421.

503 Doolittle, W. Ford. 1999. "Phylogenetic classification and the universal tree." *Science* no. 284
504 (5423):2124-2128. doi: 10.1126/science.284.5423.2124.

505 Dunning Hotopp, Julie C. 2011. "Horizontal gene transfer between bacteria and animals." *Trends in*
506 *Genetics* no. 27 (4):157-163. doi: 10.1016/j.tig.2011.01.005.

507 Eckelbarger, K. J., P. A. Tyler, and R. W. Langton. 1998. "Gonadal morphology and gametogenesis
508 in the sea pen *Pennatula aculeata* (Anthozoa: Pennatulacea) from the Gulf of Maine." *Marine*
509 *Biology* no. 132 (4):677-690.

510 Edgecombe, G. D., G. Giribet, C. W. Dunn, A. Hejnol, R. M. Kristensen, R. C. Neves, G. W. Rouse,
511 K. Worsaae, and M. V. Sorensen. 2011. "Higher-level metazoan relationships: recent progress
512 and remaining questions." *Organisms Diversity & Evolution* no. 11 (2):151-172. doi:
513 10.1007/s13127-011-0044-4.

514 Erwin, D. H., M. Laflamme, S. M. Tweedt, E. A. Sperling, D. Pisani, and K. J. Peterson. 2011. "The
515 Cambrian Conundrum: Early divergence and Later ecological success in the early history of
516 animals." *Science* no. 334 (6059):1091-1097. doi: 10.1126/science.1206375.

517 Ewen-Campen, Ben, Evelyn E. Schwager, and Cassandra G. M. Extavour. 2010. "The molecular
518 machinery of germ line specification." *Molecular Reproduction and Development* no. 77
519 (1):3-18. doi: 10.1002/mrd.21091.

520 Extavour, C. G., K. Pang, D. Q. Matus, and M. Q. Martindale. 2005. "vasa and nanos expression
521 patterns in a sea anemone and the evolution of bilaterian germ cell specification
522 mechanisms." *Evolution & Development* no. 7 (3):201-215. doi: 10.1111/j.1525-
523 142X.2005.05023.x.

524 Extavour, Cassandra G., and Michael Akam. 2003. "Mechanisms of germ cell specification across the
525 metazoans: epigenesis and preformation." *Development* no. 130 (24):5869-5884. doi:
526 10.1242/dev.00804.

527 Extavour, Cassandra G. M. 2007. "Evolution of the bilaterian germ line: lineage origin and
528 modulation of specification mechanisms." *Integrative and Comparative Biology* no. 47
529 (5):770-785. doi: 10.1093/icb/icm027.

530 Eyres, Isobel, Chiara Boschetti, Alastair Crisp, Thomas P. Smith, Diego Fontaneto, Alan Tunnacliffe,
531 and Timothy G. Barraclough. 2015. "Horizontal gene transfer in bdelloid rotifers is ancient,
532 ongoing and more frequent in species from desiccating habitats." *BMC biology* no. 13 (1):1.

533 Flot, Jean-Francois, Boris Hespeels, Xiang Li, Benjamin Noel, Irina Arkhipova, Etienne G. J.
534 Danchin, Andreas Hejnol, Bernard Henrissat, Romain Koszul, Jean-Marc Aury, Valerie
535 Barbe, Roxane-Marie Barthelemy, Jens Bast, Georgii A. Bazykin, Olivier Chabrol, Arnaud
536 Couloux, Martine Da Rocha, Corinne Da Silva, Eugene Gladyshev, Philippe Gouret, Oskar
537 Hallatschek, Bette Hecox-Lea, Karine Labadie, Benjamin Lejeune, Oliver Piskurek, Julie
538 Poulain, Fernando Rodriguez, Joseph F. Ryan, Olga A. Vakhrusheva, Eric Wajnberg,
539 Benedicte Wirth, Irina Yushenova, Manolis Kellis, Alexey S. Kondrashov, David B. Mark
540 Welch, Pierre Pontarotti, Jean Weissenbach, Patrick Wincker, Olivier Jaillon, and Karine Van
541 Doninck. 2013. "Genomic evidence for ameiotic evolution in the bdelloid rotifer *Adineta*
542 *vaga*." *Nature* no. 500 (7463):453-457.

543 Fredriksson, R., K. J. V. Nordstrom, O. Stephansson, M. G. A. Hagglund, and H. B. Schioth. 2008.
544 "The solute carrier (SLC) complement of the human genome: Phylogenetic classification
545 reveals four major families." *Febs Letters* no. 582 (27):3811-3816. doi:
546 10.1016/j.febslet.2008.10.016.

547 Funayama, N. 2010. "The stem cell system in demosponges: Insights into the origin of somatic stem
548 cells." *Development Growth & Differentiation* no. 52 (1):1-14. doi: 10.1111/j.1440-
549 169X.2009.01162.x.

550 Gaino, E., B. Burlando, and P. Buffa. 1986. "Contribution to the study of egg development and
551 derivation in *Oscarella lobularis* (Porifera, Demospongiae)." *International Journal of*
552 *Invertebrate Reproduction and Development* no. 9 (1):59-69.

553 Gaino, E., B. Burlando, L. Zunino, M. Pansini, and P. Buffa. 1984. "Origin of male gametes from
554 choanocytes in *Spongia officinalis* (Porifera, Demospongiae)." *International Journal of*
555 *Invertebrate Reproduction and Development* no. 7 (2):83-93.

556 Galliot, B., M. Mijkovic-Licina, R. de Rosa, and S. Chera. 2006. "*Hydra*, a niche for cell and
557 developmental plasticity." *Seminars in Cell & Developmental Biology* no. 17 (4):492-502.
558 doi: 10.1016/j.semcdb.2006.05.005.

559 Giani, V. C., E. Yamaguchi, M. J. Boyle, and E. C. Seaver. 2011. "Somatic and germline expression
560 of piwi during development and regeneration in the marine polychaete annelid *Capitella*
561 *teleta*." *Evodevo* no. 2:18. doi: 10.1186/2041-9139-2-10.

562 Gladyshev, Eugene A., Matthew Meselson, and Irina R. Arkhipova. 2008. "Massive horizontal gene
563 transfer in bdelloid rotifers." *Science* no. 320 (5880):1210-1213. doi:
564 10.1126/science.1156407.

565 Gogarten, J. Peter, W. Ford Doolittle, and Jeffrey G. Lawrence. 2002. "Prokaryotic evolution in light
566 of gene transfer." *Molecular Biology and Evolution* no. 19 (12):2226-2238.

567 Grant, Jessica R., and Laura A. Katz. 2014. "Building a phylogenomic pipeline for the eukaryotic tree
568 of life - addressing deep phylogenies with genome-scale data." *PLoS Currents* no. 6. doi:
569 10.1371/currents.tol.c24b6054aebf3602748ac042ccc8f2e9.

570 Gregory, T. Ryan. 2001. "Coincidence, coevolution, or causation? DNA content, cellsize, and the C-
571 value enigma." *Biological Reviews* no. 76 (1):65-101. doi: 10.1111/j.1469-
572 185X.2000.tb00059.x.

573 Gregory, T. Ryan. 2016. Animal Genome Size Database. http://www.genomesize.com.

574 Haegeman, Annelies, John T. Jones, and Etienne G. J. Danchin. 2011. "Horizontal gene transfer in
575 nematodes: a catalyst for plant parasitism?" *Molecular Plant-Microbe Interactions* no. 24
576 (8):879-887. doi: 10.1094/MPMI-03-11-0055.

577 Harbison, G. R., and R. L. Miller. 1986. "Not all ctenophores are hermaphrodites. Studies on the
578 systematics, distribution, sexuality and development of two species of *Ocyropsis*." *Marine*
579 *Biology* no. 90 (3):413-424.

580 Huang, Jinling. 2013. "Horizontal gene transfer in eukaryotes: The weak-link model." *BioEssays* no.
581 35 (10):868-875. doi: 10.1002/bies.201300007.

582 Jackson, Daniel J., Luciana Macis, Joachim Reitner, and Gert Wörheide. 2011. "A horizontal gene
583 transfer supported the evolution of an early metazoan biomineralization strategy." *BMC*
584 *Evolutionary Biology* no. 11 (1). doi: 10.1186/1471-2148-11-238.

585 Jain, Ravi, Maria C. Rivera, and James A. Lake. 1999. "Horizontal gene transfer among genomes: the
586 complexity hypothesis." *Proceedings of the National Academy of Sciences of the United*
587 *States of America* no. 96 (7):3801-3806.

588 Jain, Ravi, Maria C. Rivera, Jonathan E. Moore, and James A. Lake. 2003. "Horizontal gene transfer
589 accelerates genome innovation and evolution." *Molecular Biology and Evolution* no. 20
590 (10):1598-1602. doi: 10.1093/molbev/msg154.

591 Juliano, C. E., S. Z. Swartz, and G. M. Wessel. 2010. "A conserved germline multipotency program."
592 *Development* no. 137 (24):4113-4126. doi: 10.1242/dev.047969.

593 Just, J., R. M. Kristensen, and J. Olesen. 2014. "Dendrogramma, new genus, with two new non-
594 bilaterian species from the marine bathyal of southeastern Australia (animalia, metazoa
595 incertae sedis) - with similarities to some medusoids from the precambrian ediacara." *Plos*
596 *One* no. 9 (9). doi: 10.1371/journal.pone.0102976.

597 Kanehisa, Minoru, and Susumu Goto. 2000. "KEGG: kyoto encyclopedia of genes and genomes."
598 *Nucleic acids research* no. 28 (1):27-30.

599 Katz, Laura A. 2002. "Lateral gene transfers and the evolution of eukaryotes: theories and data."
600 *International Journal of Systematic and Evolutionary Microbiology* no. 52 (5):1893-1900.
601 doi: 10.1099/ijs.0.02113-0.

602 Katz, Laura A. 2012. "Origin and diversification of eukaryotes." *Annual review of microbiology* no.
603 66:411-427.

604 Katz, Laura A. 2015. "Recent events dominate interdomain lateral gene transfers between prokaryotes
605 and eukaryotes and, with the exception of endosymbiotic gene transfers, few ancient transfer
606 events persist." *Phil. Trans. R. Soc. B* no. 370 (1678):20140324.

607 Keeling, Patrick J., and Jeffrey D. Palmer. 2008. "Horizontal gene transfer in eukaryotic evolution."
608 *Nature Reviews Genetics* no. 9 (8):605-618. doi: 10.1038/nrg2386.

609 Klasson, Lisa, Zakaria Kambris, Peter E. Cook, Thomas Walker, and Steven P. Sinkins. 2009.
610 "Horizontal gene transfer between *Wolbachia* and the mosquito *Aedes aegypti*." *BMC*
611 *Genomics* no. 10 (1). doi: 10.1186/1471-2164-10-33.

612 Komai, Taku. 1922. Studies on two aberrant ctenophores: Coeloplana and Gastrodes. Kyoto: Self-
613 Published.

614 Kondrashov, Fyodor A., Eugene V. Koonin, Igor G. Morgunov, Tatiana V. Finogenova, and Marie N.
615 Kondrashova. 2006. "Evolution of glyoxylate cycle enzymes in Metazoa: evidence of

multiple horizontal transfer events and pseudogene formation." *Biology Direct* no. 1 (1):1-14. doi: 10.1186/1745-6150-1-31.

Koning, Audrey P. de, Fiona S. L. Brinkman, Steven J. M. Jones, and Patrick J. Keeling. 2000. "Lateral gene transfer and metabolic adaptation in the human parasite *Trichomonas vaginalis*." *Molecular Biology and Evolution* no. 17 (11):1769-1773.

Lawrence, Jeffrey G., and Heather Hendrickson. 2003. "Lateral gene transfer: when will adolescence end?" *Molecular Microbiology* no. 50 (3):739-749.

Lawrence, Jeffrey G., and Howard Ochman. 2002. "Reconciling the many faces of lateral gene transfer." *Trends in Microbiology* no. 10 (1):1-4.

Lehmann, Ruth, and Brian E. Richardson. 2010. "Mechanisms guiding primordial germ cell migration: strategies from different organisms." *Nature Reviews Molecular Cell Biology* no. 11 (1).

Lynch, M., and J. S. Conery. 2000. "The evolutionary fate and consequences of duplicate genes." *Science* no. 290 (5494):1151-1155. doi: 10.1126/science.290.5494.1151.

Nakashima, Keisuke, Lixy Yamada, Yutaka Satou, Jun-ichi Azuma, and Nori Satoh. 2004. "The evolutionary origin of animal cellulose synthase." *Development Genes and Evolution* no. 214 (2):81-88. doi: 10.1007/s00427-003-0379-8.

Nelson, Karen E., Rebecca A. Clayton, Steven R. Gill, Michelle L. Gwinn, Robert J. Dodson, Daniel H. Haft, Erin K. Hickey, Jeremy D. Peterson, William C. Nelson, Karen A. Ketchum, Lisa McDonald, Teresa R. Utterback, Joel A. Malek, Katja D. Linher, Mina M. Garrett, Ashley M. Stewart, Matthew D. Cotton, Matthew S. Pratt, Cheryl A. Phillips, Delwood Richardson, John Heidelberg, Granger G. Sutton, Robert D. Fleischmann, Jonathan A. Eisen, Owen White, Steven L. Salzberg, Hamilton O. Smith, J. Craig Venter, and Claire M. Fraser. 1999.

"Evidence for lateral gene transfer between Archaea and Bacteria from genome sequence of *Thermotoga maritima*." *Nature* no. 399 (6734):323-329. doi: 10.1038/20601.

Newmark, P. A., Y. Wang, and T. Chong. 2008. "Germ cell specification and regeneration in planarians." *Control and Regulation of Stem Cells* no. 73:573-581. doi: 10.1101/sqb.2008.73.022.

Nikoh, Naruo, and Atsushi Nakabachi. 2009. "Aphids acquired symbiotic genes via lateral gene transfer." *BMC Biology* no. 7 (1). doi: 10.1186/1741-7007-7-12.

Nishimiya-Fujisawa, C., and S. Kobayashi. 2012. "Germline stem cells and sex determination in *Hydra*." *International Journal of Developmental Biology* no. 56 (6-8):499-508. doi: 10.1387/ijdb.123509cf.

Ochman, Howard, Jeffrey G. Lawrence, and Eduardo A. Groisman. 2000. "Lateral gene transfer and the nature of bacterial innovation." *Nature* no. 405 (6784):299-304. doi: 10.1038/35012500.

Okamoto, M., H. Ohsawa, T. Hayashi, K. Owaribe, and P. A. Tsonis. 2007. "Regeneration of retinotectal projections after optic tectum removal in adult newts." *Molecular Vision* no. 13 (239-41):2112-2118.

Paganini, Julien, Amandine Campan-Fournier, Martine Da Rocha, Philippe Gouret, Pierre Pontarotti, Eric Wajnberg, Pierre Abad, and Etienne G. J. Danchin. 2012. "Contribution of lateral gene transfers to the genome composition and parasitic ability of root-knot nematodes." *PLoS One* no. 7 (11):e50875.

Pál, Csaba, Balázs Papp, and Martin J. Lercher. 2005. "Adaptive evolution of bacterial metabolic networks by horizontal gene transfer." *Nature Genetics* no. 37 (12):1372-1375. doi: 10.1038/ng1686.

661 Penn, Osnat, Eyal Privman, Giddy Landan, Dan Graur, and Tal Pupko. 2010. "An alignment
662 confidence score capturing robustness to guide tree uncertainty." *Molecular Biology and*
663 *Evolution* no. 27 (8):1759-1767. doi: 10.1093/molbev/msq066.

664 Price, M. N., P. S. Dehal, and A. P. Arkin. 2009. "FastTree: Computing large minimum evolution
665 trees with profiles instead of a distance matrix." *Molecular Biology and Evolution* no. 26
666 (7):1641-1650. doi: 10.1093/molbev/msp077.

667 Price, M. N., P. S. Dehal, and A. P. Arkin. 2010. "FastTree 2: approximately maximum-likelihood
668 trees for large alignments." *Plos One* no. 5 (3):10. doi: 10.1371/journal.pone.0009490.

669 Ragan, Mark A. 2001. "On surrogate methods for detecting lateral gene transfer." *FEMS*
670 *Microbiology letters* no. 201 (2):187-191.

671 Ragan, Mark A., Timothy J. Harlow, and Robert G. Beiko. 2006. "Do different surrogate methods
672 detect lateral genetic transfer events of different relative ages?" *Trends in microbiology* no. 14
673 (1):4-8.

674 Ryan, J. F., K. Pang, C. E. Schnitzler, A. D. Nguyen, R. T. Moreland, D. K. Simmons, B. J. Koch, W.
675 R. Francis, P. Havlak, S. A. Smith, N. H. Putnam, S. H. D. Haddock, C. W. Dunn, T. G.
676 Wolfsberg, J. C. Mullikin, M. Q. Martindale, A. D. Baxevanis, and Nisc Comparative
677 Sequencing Progra. 2013. "The Genome of the Ctenophore *Mnemiopsis leidyi* and its
678 Implications for cell type evolution." *Science* no. 342 (6164):1336-+. doi:
679 10.1126/science.1242592.

680 Schönknecht, Gerald, Wei-Hua Chen, Chad M. Ternes, Guillaume G. Barbier, Roshan P. Shrestha,
681 Mario Stanke, Andrea Bräutigam, Brett J. Baker, Jillian F. Banfield, and R. Michael Garavito.
682 2013. "Gene transfer from bacteria and archaea facilitated evolution of an extremophilic
683 eukaryote." *Science* no. 339 (6124):1207-1210.

684 Schönknecht, Gerald, Andreas P. M. Weber, and Martin J. Lercher. 2014. "Horizontal gene
685 acquisitions by eukaryotes as drivers of adaptive evolution." *BioEssays* no. 36 (1):9-20. doi:
686 10.1002/bies.201300095.

687 Shibata, T. F., T. Oji, K. Akasaka, and K. Agata. 2010. "Staging of regeneration process of an arm of
688 the feather star *Oxycomanthus japonicus* focusing on the oral-aboral boundary."
689 *Developmental Dynamics* no. 239 (11):2947-2961. doi: 10.1002/dvdy.22429.

690 Shu, D. G., Y. Isozaki, X. L. Zhang, J. Han, and S. Maruyama. 2014. "Birth and early evolution of
691 metazoans." *Gondwana Research* no. 25 (3):884-895. doi: 10.1016/j.gr.2013.09.001.

692 Simpson, T. L. 1980. "Reproductive processes in sponges - a critical-evaluation of current data and
693 views." *International Journal of Invertebrate Reproduction* no. 2 (4):251-269.

694 Simpson, T. L. 1984. "The cell biology of sponges." *The cell biology of sponges.*:i-xiv, 1-662.

695 Sloan, Daniel B., and Nancy A. Moran. 2012. "Endosymbiotic bacteria as a source of carotenoids in
696 whiteflies." *Biology letters* no. 8 (6):986-989.

697 Soucy, Shannon M., Jinling Huang, and Johann Peter Gogarten. 2015. "Horizontal gene transfer:
698 building the web of life." *Nature Reviews Genetics* no. 16 (8):472-482.

699 Starcevic, Antonio, Shamima Akthar, Walter C. Dunlap, J. Malcolm Shick, Daslav Hranueli, John
700 Cullum, and Paul F. Long. 2008. "Enzymes of the shikimic acid pathway encoded in the
701 genome of a basal metazoan, *Nematostella vectensis*, have microbial origins." *Proceedings of*
702 *the National Academy of Sciences* no. 105 (7):2533-2537. doi: 10.1073/pnas.0707388105.

703 Starz-Gaiano, Michelle, and Ruth Lehmann. 2001. "Moving towards the next generation."
704 *Mechanisms of Development* no. 105 (1–2):5-18. doi: 10.1016/S0925-4773(01)00392-6.

705 Syvanen, Michael. 2012. "Evolutionary implications of horizontal gene transfer." *Annual Review of*
706 *Genetics* no. 46 (1):341-358. doi: 10.1146/annurev-genet-110711-155529.

707 Tsurumi, M., and H. M. Reiswig. 1997. "Sexual versus asexual reproduction in an oviparous rope-
708 form sponge, *Aplysina cauliformis* (Porifera; Verongida)." *Invertebrate Reproduction &*
709 *Development* no. 32 (1):1-9. doi: 10.1080/07924259.1997.9672598.

710 Wedi, Steven E., and Daphne Fautin Dunn. 1983. "Gametogenesis and reproductive periodicity of the
711 subtidal sea anemone *Urticina lofotensis* (Coelenterata: Actiniaria) in California." *The*
712 *Biological Bulletin* no. 165 (2):458-472.

713 Whitaker, John W., Glenn A. McConkey, and David R. Westhead. 2009. "The transferome of
714 metabolic genes explored: analysis of the horizontal transfer of enzyme encoding genes in
715 unicellular eukaryotes." *Genome Biology* no. 10 (4):1-13. doi: 10.1186/gb-2009-10-4-r36.

716 Wijayawardena, Bhagya K., Dennis J. Minchella, and J. Andrew DeWoody. 2013. "Hosts, parasites,
717 and horizontal gene transfer." *Trends in parasitology* no. 29 (7):329-338.

718 Wolf, Yuri I., and Eugene V. Koonin. 2013. "Genome reduction as the dominant mode of evolution."
719 *Bioessays* no. 35 (9):829-837.

720 Zhang, Z. Q. 2013. "Animal biodiversity: An update of classification and diversity in 2013." *Zootaxa*
721 no. 3703 (1):5-11.

722

Figure 1: Example of a gene present only in Metazoa and bacteria/archaea

This single-gene tree is for the gene URAD, a gene involved in a uric acid degradation pathway, which was transferred during event D (see Table 1, S2, and Figure 3). The phylogeny was built from the URAD gene sequences that were present in any of the 910 taxa in our pipeline, serving as an example of a gene present only in Metazoa and bacteria/archaea (e.g. presence/absence data). Colors corresponding to major taxonomic groups are labeled in the bottom right corner.

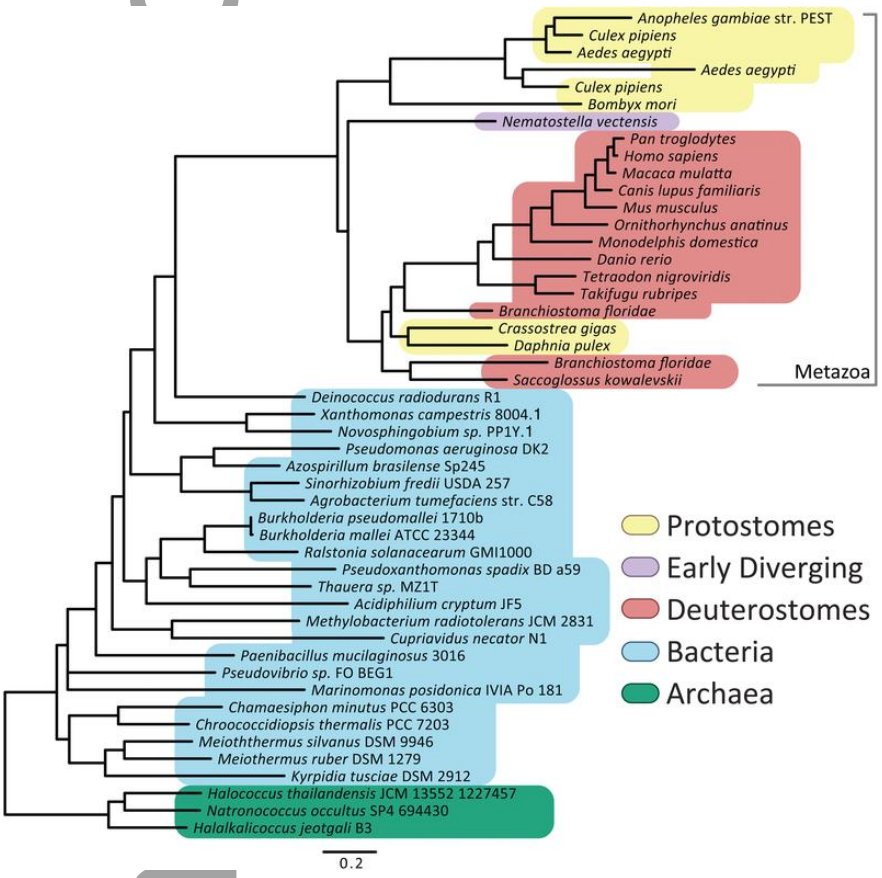


Figure 2: Tree topology trends in genes laterally transferred twice

Diagrams summarizing the two trends found in the topology of genes laterally transferred twice. **A)** Describes genes that were also transferred into one other eukaryotic supergroup in addition to a transfer into Metazoa. **B)** These topologies showed EGT in addition to the LGT from bacteria/archaea into Metazoa. See text for further details.

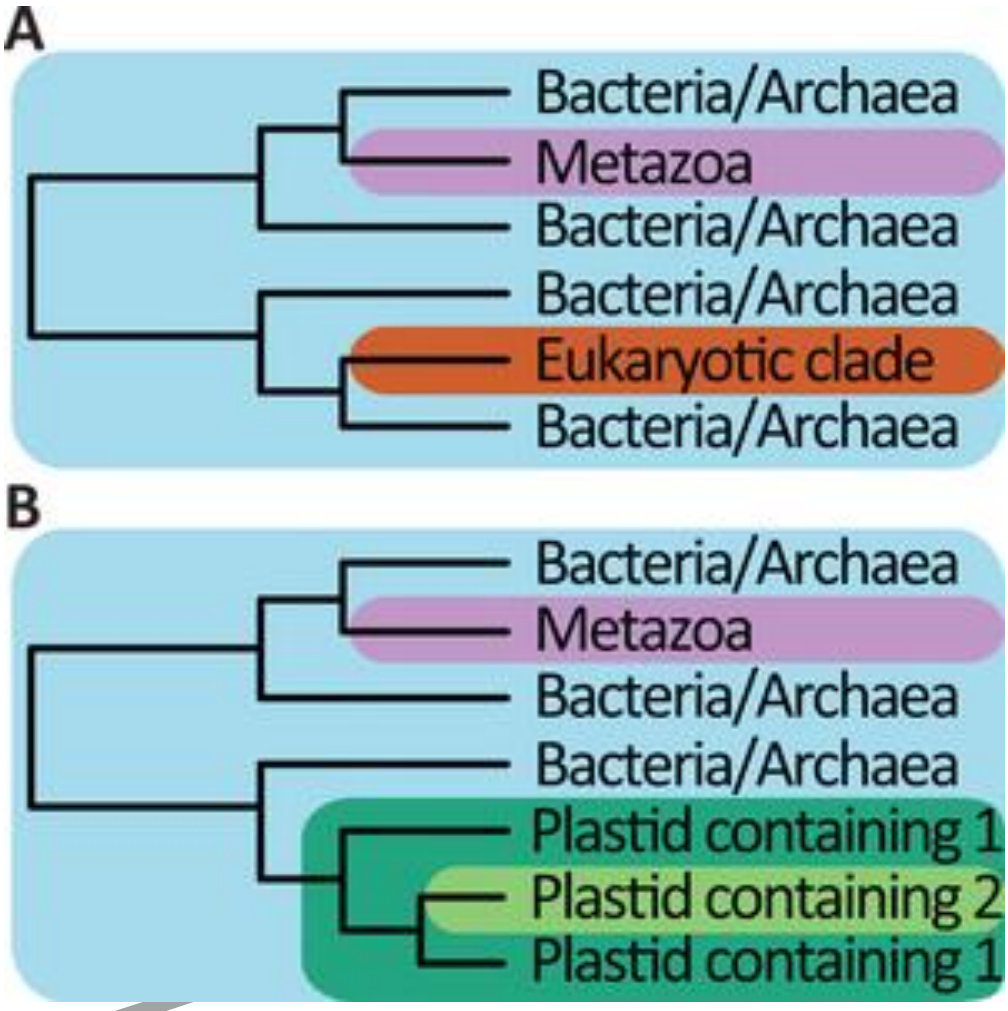
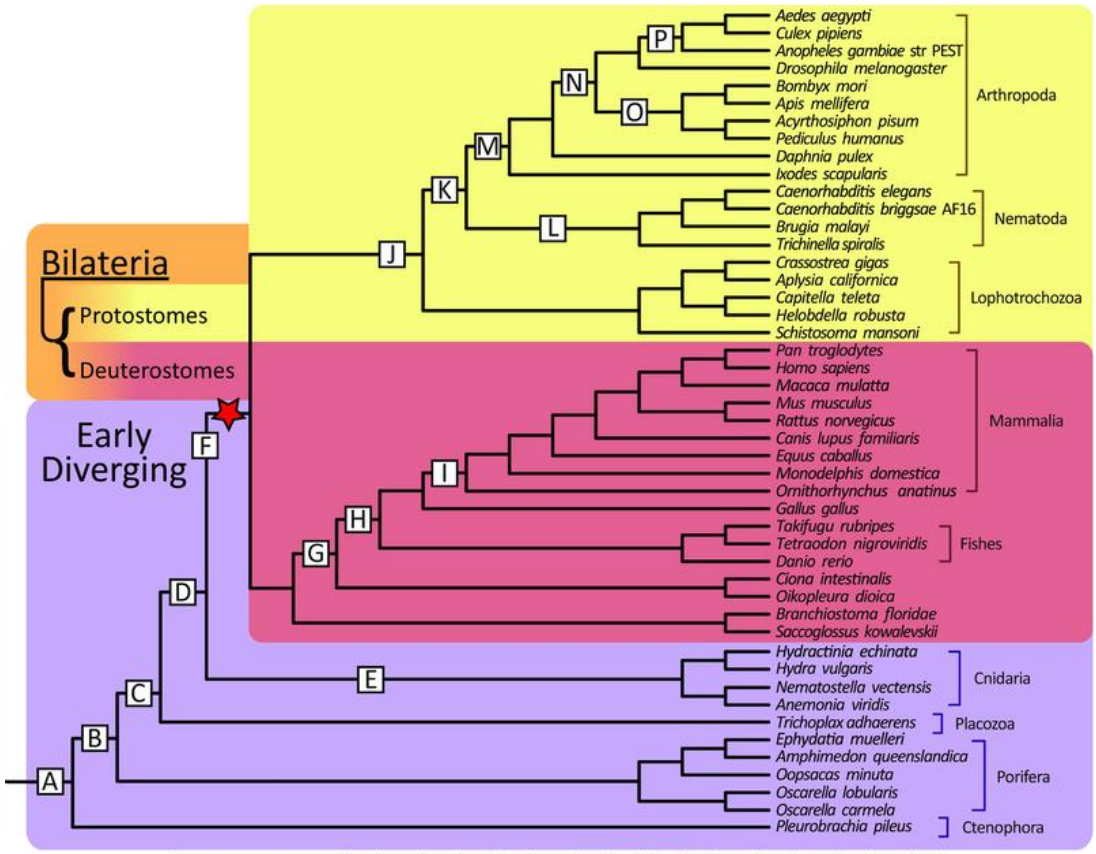


Figure 3: Estimate of LGTs across the metazoan tree of life

We charted LGTs onto a synthetic tree of metazoan evolution to approximate the time of transfer for each gene analyzed in this study. The tree includes only the Metazoa whose genomes contained one of the 58 laterally transferred genes found in this study (see table S1 for a complete list of metazoan taxa in the pipeline). The letters in the square boxes are categories that bin the LGT events found in this study by the approximate time of transfer. The table at the bottom provides a numerical summary of how many genes were found per event, and the letters correspond to Table 1 as well as Table S2 for more detailed information on the genes within that event. The star denotes when a sequestered germline developed. See methods for detailed information on how this synthetic tree was created. Branch lengths are not proportional to time.



Transfer Event	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
# of Genes	1	7	4	13	2	7	1	7	2	1	2	2	1	6	1	1

746

747